



Dual-model analysis for improving the discrimination performance of human and nonhuman blood based on Raman spectroscopy

HAIYI BIAN,^{1,2} PENG WANG,¹ NING WANG,¹ YUBING TIAN,¹ PENGLI BAI,³
HAOWEN JIANG,⁴ AND JING GAO^{1,*}

¹Jiangsu Key Laboratory of Medical Optics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China

²Schott Glass Technologies (Suzhou) Co., Ltd., Suzhou 215009, China

³CAS Key Lab of Bio-Medical Diagnostics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Jiangsu 215163, China

⁴Department of Urology, Huashan Hospital, Fudan University, Shanghai 200040, China

*owengaojing@126.com

Abstract: The discrimination accuracy for human and nonhuman blood is important for customs inspection and forensic applications. Recently, Raman spectroscopy has shown effectiveness in analyzing blood droplets and stains with an excitation wavelength of 785 nm. However, the discrimination of liquid whole blood in a vacuum blood tube using Raman spectroscopy, which is a form of noncontact and nondestructive detection, has not been achieved. An excitation wavelength of 532 nm was chosen to avoid the fluorescent background of the blood tube, at the cost of reduced spectroscopic information and discrimination accuracy. To improve the accuracy and true positive rate (TPR) for human blood, a dual-model analysis method is proposed. First, model 1 was used to discriminate human-unlike nonhuman blood. Model 2 was then used to discriminate human-like nonhuman blood from the “human blood” obtained by model 1. A total of 332 Raman spectra from 10 species were used to build and validate the model. A blind test and external validation demonstrated the effectiveness of the model. Compared with the results obtained by the single partial least-squares model, the discrimination performance was improved. The total accuracy and TPR, which are highly important for practical applications, increased to 99.1% and 97.4% from 87.2% and 90.6%, respectively.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (180.5655) Raman microscopy; (300.6450) Spectroscopy, Raman; (200.3050) Information processing

References and links

1. X. Wan, J. Wang, P. X. Liu, and T. T. Zhang, “Identification of animal whole blood based on near infrared transmission spectroscopy,” *Guangpuxue Yu Guangpu Fenxi* **36**(1), 80–83 (2016).
2. N. Suwa, H. Ikegaya, T. Takasaka, K. Nishigaki, and K. Sakurada, “Human blood identification using the genome profiling method,” *Leg. Med. (Tokyo)* **14**(3), 121–125 (2012).
3. L. Zhang, M. Zhou, X. Li, G. Li, and L. Lin, “Discrimination of human and nonhuman blood using visible diffuse reflectance spectroscopy,” *Anal. Methods-UK* **6**(23), 9419–9423 (2014).
4. K. Virkler and I. K. Lednev, “Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene,” *Forensic Sci. Int.* **188**(1-3), 1–17 (2009).
5. Y. Y. Shen, X. Chen, and R. W. Murphy, “Assessing DNA barcoding as a tool for species identification and data quality control,” *PLoS One* **8**(2), e57125 (2013).
6. D. L. Dalton and A. Kotze, “DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa,” *Forensic Sci. Int.* **207**(1-3), e51–e54 (2011).
7. I. Aly, R. Zalat, B. W. E. I. Aswad, I. M. Moharm, B. M. Masound, and T. Diab, “Novel nanomagnetic beads based–latex agglutination assay for rapid diagnosis of human schistosomiasis haematobium,” *Int. J. Med. Health Biomed. Pharm. Eng.* **7**(12), 642–647 (2013).
8. H. Inoue, F. Takabe, O. Takenaka, M. Iwasa, and Y. Maeno, “Species identification of blood and bloodstains by high-performance liquid chromatography,” *Int. J. Legal Med.* **104**(1), 9–12 (1990).

9. O. Goldenberg, S. Herrmann, T. Adam, G. Marjoram, G. Hong, U. B. Göbel, and B. Graf, "Use of denaturing high-performance liquid chromatography for rapid detection and identification of seven *Candida* species," *J. Clin. Microbiol.* **43**(12), 5912–5915 (2005).
10. E. O. Espinoza, M. A. Kirms, and M. S. Filipek, "Identification and quantitation of source from hemoglobin of blood and blood mixtures by high performance liquid chromatography," *J. Forensic Sci.* **41**(5), 804–811 (1996).
11. E. O. Espinoza, N. C. Lindley, K. M. Gordon, J. A. Ekhooff, and M. A. Kirms, "Electrospray ionization mass spectrometric analysis of blood for differentiation of species," *Anal. Biochem.* **268**(2), 252–261 (1999).
12. K. M. Elkins, "Rapid presumptive "Fingerprinting" of Body Fluids and Materials by ATR FT-IR Spectroscopy," *J. Forensic Sci.* **56**(6), 1580–1587 (2011).
13. E. Mistek and I. K. Lednev, "Identification of species' blood by attenuated total reflection (ATR) Fourier transform infrared (FT-IR) spectroscopy," *Anal. Bioanal. Chem.* **407**(24), 7435–7442 (2015).
14. L. Zhang, S. Zhang, M. Sun, Z. Wang, H. Li, Y. Li, G. Li, and L. Lin, "Blood species identification using Near-Infrared diffuse transmitted spectra and PLS-DA method," *Infrared Phys. Technol.* **76**, 587–591 (2016).
15. L. Zhang, M. Sun, Z. Wang, H. Li, Y. Li, Z. Fu, Y. Guan, G. Li, and L. Lin, "Optimal wavelength selection for visible diffuse reflectance spectroscopy discriminating human and nonhuman blood species," *Anal. Methods-UK* **8**(2), 381–385 (2016).
16. G. McLaughlin, K. C. Doty, and I. K. Lednev, "Discrimination of human and animal blood traces via Raman spectroscopy," *Forensic Sci. Int.* **238**, 91–95 (2014).
17. J. Fujihara, Y. Fujita, T. Yamamoto, N. Nishimoto, K. Kimura-Kataoka, S. Kurata, Y. Takinami, T. Yasuda, and H. Takeshita, "Blood identification and discrimination between human and nonhuman blood using portable Raman spectroscopy," *Int. J. Legal Med.* **131**(2), 319–322 (2017).
18. H. Bian, P. Wang, J. Wang, H. Yin, Y. Tian, P. Bai, X. Wu, N. Wang, Y. Tang, and J. Gao, "Discrimination of human and nonhuman blood using Raman spectroscopy with self-reference algorithm," *J. Biomed. Opt.* **22**(9), 1–7 (2017).
19. K. Virkler and I. K. Lednev, "Blood species identification for forensic purposes using Raman spectroscopy combined with advanced statistical analysis," *Anal. Chem.* **81**(18), 7773–7777 (2009).
20. G. McLaughlin, K. C. Doty, and I. K. Lednev, "Raman spectroscopy of blood for species identification," *Anal. Chem.* **86**(23), 11628–11633 (2014).
21. P. Bai, J. Wang, H. Yin, Y. Tian, W. Yao, and J. Gao, "Discrimination of Human and Nonhuman Blood by Raman Spectroscopy and Partial Least Squares Discriminant Analysis," *Anal. Lett.* **50**(2), 379–388 (2017).
22. K. C. Doty and I. K. Lednev, "Differentiation of human blood from animal blood using Raman spectroscopy: A survey of forensically relevant species," *Forensic Sci. Int.* **282**, 204–210 (2018).
23. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.* **46**(1/3), 389–422 (2002).
24. L. Zhang, S. Zhang, M. Sun, H. Li, Y. Li, Z. Fu, Y. Guan, G. Li, and L. Lin, "Noncontact discrimination of animal and human blood with vacuum blood vessel and factors affect the discrimination," *Infrared Phys. Technol.* **81**, 210–214 (2017).
25. H. Bian and J. Gao, "Error analysis of the spectral shift for partial least squares models in Raman spectroscopy," *Opt. Express* **26**(7), 8016–8027 (2018).
26. Z. M. Zhang, S. Chen, and Y. Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst (Lond.)* **135**(5), 1138–1146 (2010).
27. S. J. Baek, A. Park, Y. J. Ahn, and J. Choo, "Baseline correction using asymmetrically reweighted penalized least squares smoothing," *Analyst (Lond.)* **140**(1), 250–257 (2015).

1. Introduction

Blood carries genetic information. Moreover, blood from people of different nationalities carries different genetic information. Using the genetic information of whole blood, biological weapons against a particular nationality may be developed by terrorists. On the other hand, genetic information from blood can help identify a victim or suspect and solve a criminal case. To prevent genetic terrorist activities and combat crime, correctly discriminating human blood is important for customs inspection departments and forensic investigations [1–4]. Many methods, such as, DNA tests [5, 6], latex agglutination [7], high-performance liquid chromatography [8–10], and mass spectrometry [11], have been used to discriminate human and nonhuman blood and have exhibited effectiveness. However, these methods require reagents and are time-consuming and destructive to samples. Moreover, they are dangerous for practitioners because of the pretreatment of the blood samples.

Spectral analysis technology combined with chemometric algorithms has recently attracted attention for the identification of the species of origin of blood samples. Attenuated total reflection Fourier transform infrared spectroscopy (ATR FT-IR) was first shown to be effective in a body-fluid examination by Elkins [12]. Subsequently, the effectiveness of species differentiation according to dry traces of blood using ATR FT-IR was demonstrated

by Lednev [13]. Dry blood samples, which are typically found at a crime scene, were used to demonstrate the method. Blood from humans, cats, and dogs were analyzed, and 100% accuracy was achieved. Near-infrared transmitted spectroscopy was synchronously demonstrated by Zhang [1] and Lin [14] to be effective for discriminating human and nonhuman blood. In 2014, a method combining visible reflectance spectroscopy with partial least-squares discriminant analysis (PLS-DA) was demonstrated by Lin to be a powerful tool for species differentiation [15]. Moreover, Raman spectroscopy was reported as an effective technology for discriminating human and nonhuman blood [16–22]. Given its advantages—such as, being nondestructive, not requiring sample preparation, and having high accuracy for the identification of organic, inorganic, and biological species—Raman spectroscopy has shown greater potential for species identification applications than other spectral methods. Because the compositions of human and nonhuman blood are similar, the Raman peaks for both types of blood samples are the same. Chemometric algorithms, including principal component analysis [18], PLS-DA [19, 22], support vector machines [23], and self-reference algorithms [18], are used to identify the species of origin. All the aforementioned methods were proven to be effective for blood samples, which should be transferred from the vacuum tube onto a slide coated with aluminum by practitioners. However, this sampling process may expose the practitioners to infection risks. Thus, considering the safety of the practitioners, we explored a discrimination method involving measurement of blood samples in their original containers (vacuum blood tubes), which has been considered in diffuse reflectance spectroscopy [24]. Thus far, the discrimination of human and nonhuman blood with the samples in the vacuum blood tubes via Raman spectroscopy has not been proven by other groups. This is possibly because the fluorescence of the vacuum blood tube is so high that the Raman spectra of the blood disappear at an excitation wavelength of 785 nm, which has been demonstrated as the most suitable excitation wavelength for blood samples, as the fluorescence of blood is minimized. To avoid the influence of the fluorescence background of the vacuum blood tube, the excitation wavelength of 532 nm is used, and the Raman spectra of the blood in the vacuum blood tube are obtained. Analysis of the Raman spectra reveals that the identification accuracy of the PLS model is lower than that of Ref [22]. (in this paper, the excitation wavelength is 785 nm) owing to the small difference in the Raman spectra of nonhuman blood from different species.

To solve this problem, in this study, we proposed a dual-model analysis method, in which the first model was used to discriminate human and nonhuman blood for all blood samples, and the second model was used to discriminate human and nonhuman blood using the “human blood” determined by the first model. The nonhuman blood samples differed between the training data sets used to build the two models. For the first model, the nonhuman-blood training data were those whose Raman spectra were not similar to those of human blood. As a result, the discrimination of human blood or human-like nonhuman blood and nonhuman blood was achieved. The human blood or human-like nonhuman blood determined by the first model was then processed using the second model. The training data set for the second model comprised human blood and human-like nonhuman blood, while that for the first model comprised the same human blood and human-unlike nonhuman blood. Compared with the results obtained by the traditional PLS model, the discrimination accuracy and especially the true positive rate (TPR) of human blood are improved.

2. Method and system

2.1 System

The Raman spectroscopy used in this work was similar to that in our previous work [25]. The optical system comprised three parts: a light source, a microscope, and a spectrograph. The wavelength of the light source (SLM G-100, Diode Tek) was 532 nm. A CX40M (Sunny Optical Technology, China) microscope was used to illuminate the sample and collect the Raman signal. The collected Raman signal was transmitted to a spectrograph (SR-303i-B,

Andor) through a fiber. The spectral resolution of the spectrograph was $\sim 2.6 \text{ cm}^{-1}$. To avoid photodegradation and sample damage, the laser power incident on the blood was $\sim 3 \text{ mW}$. The laser exposure time was 1 s, and the accumulation was 7.

2.2 Materials

Whole blood samples from a total of 159 human donors and 173 animal donors, including chickens, ducks, geese, doves, pigs, and sheep, were used. All the animal donors were provided by the Laboratory Animal Center of Soochow University. Before the blood was collected, all the donors were ensured to be healthy and not be taking medication. All experiments were performed in compliance with the law. The vacuum tube, which was made of glass, contained dipotassium ethylenediaminetetraacetic acid ($\text{K}_2 \text{ EDTA}$) at a concentration of $\sim 1.6 \text{ mg K}_2 \text{ EDTA per 1 mL}$ of whole blood. The blood samples were immediately preserved in an icebox after they were collected. The spectra were obtained between 24 and 72 h after the blood samples were acquired. The preservation time of the blood was demonstrated to be as long as 3 months [17]. The thickness of the vacuum blood tube was $\sim 1 \text{ mm}$; thus, a long-working distance objective lens (N PLAN H, $50\times$, 0.5 NA , 7.1-mm working distance, Leica) was used to focus the light onto the surface of the blood in the tube. A total of 332 Raman spectra were collected, including 159 Raman spectra of human blood, 26 Raman spectra of chicken blood, 29 Raman spectra of duck blood, 32 Raman spectra of goose blood, 27 Raman spectra of dove blood, 12 Raman spectra of pig blood, 21 Raman spectra of sheep blood, 16 Raman spectra of rabbit blood, five Raman spectra of mouse blood, and five Raman spectra of dog blood. Each spectrum was collected from one donor. To ensure diversity, the ages of the human donors varied (number of donors aged between 18 and 34: 38; number of donors aged between 35 and 60: 74; number of donors aged above 60: 47), and blood samples from both males and females (number of males: 78; number of females: 81) were used. The race of the human donors was Chinese. The blood samples were mixed before the Raman spectra were measured.

2.3 Model

The Raman spectra of the blood from different species are presented in Fig. 1. Figure 1(a) shows the original spectra, and Fig. 1(b) shows the preprocessed spectra with subtraction of the background and normalization. The background of the Raman spectra was removed using the airPLS method [26, 27], and the Raman spectra were normalized by dividing the maximal value of the spectra by the background subtracted. In the figure, the main Raman peaks of human and nonhuman blood are similar. According to the literature, no obvious difference is observed among the Raman spectra of nonhuman blood collected from different species under 785-nm laser excitation [18, 20]. However, when the excitation wavelength was 532 nm, the Raman spectra of the blood collected from chickens, ducks, geese, and doves differed from those collected from pigs, rabbits, and sheep. As shown in Fig. 2, the Raman spectra of the blood collected from pigs, rabbits, and sheep are similar to those for humans. To improve the discrimination accuracy, we divided the blood samples into three groups: 1) human blood, 2) human-unlike nonhuman blood, and 3) human-like nonhuman blood. Accordingly, the traditional PLS model used for discriminating blood was separated into two models. Figure 2 illustrates the analysis procedures for the traditional single-model method and the proposed dual-model method. The final discrimination of human blood was performed using model 2, whereas the discrimination of nonhuman blood involved the comprehensive consideration of the discrimination of nonhuman blood achieved by models 1 and 2. The training set was used to construct the PLS model with six latent variables, and the leave-one-out method was used for internal cross-validation. For the models, the values 1 and 2 represent human and nonhuman blood, respectively, and the threshold is set as 1.5.

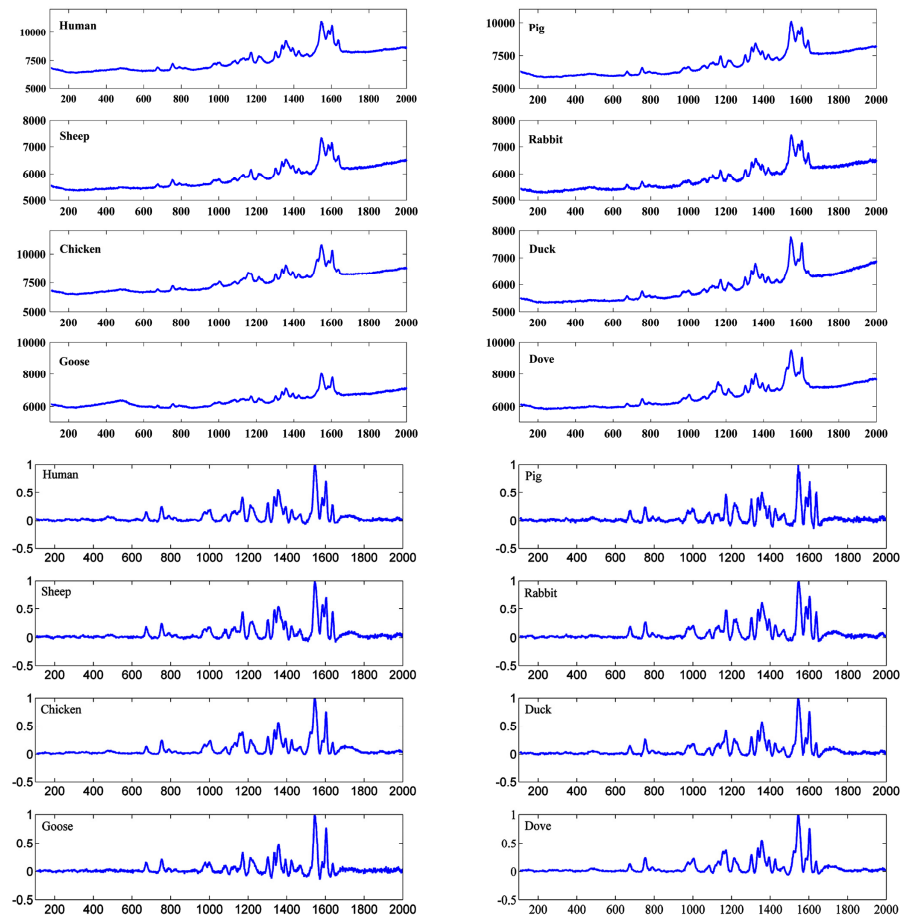


Fig. 1. Raman spectra of blood originating from different species.

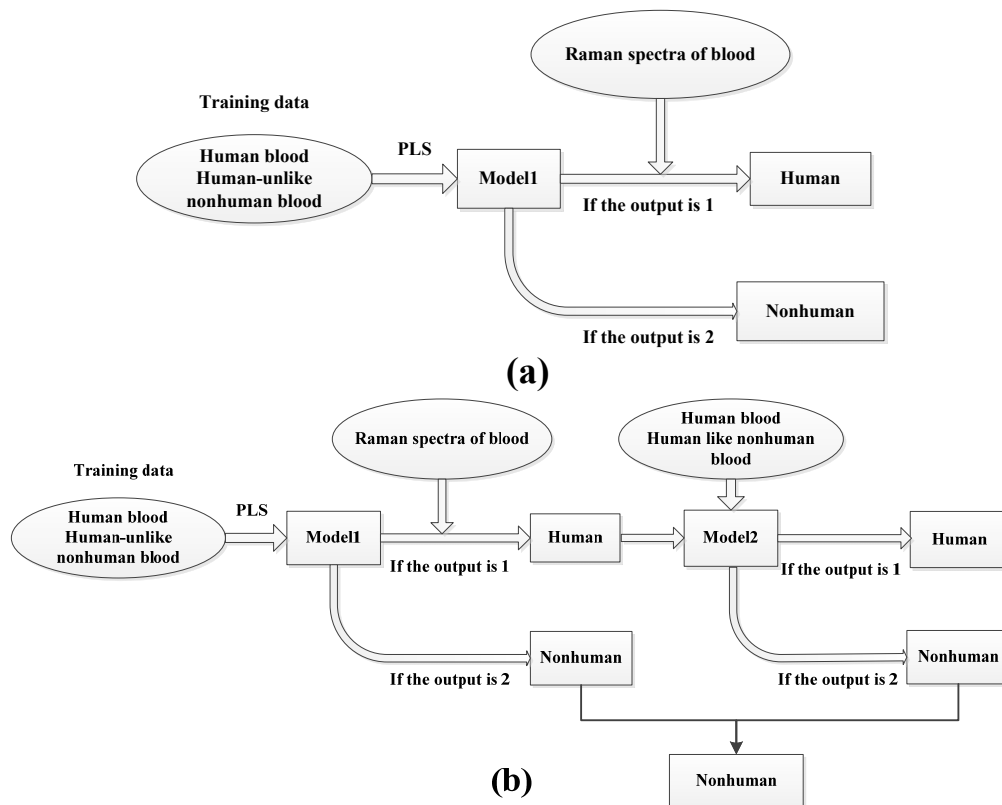


Fig. 2. Flow charts of the (a) single-model and (b) dual-model methods.

3. Result

3.1 Prediction value of prediction set obtained via dual-model analysis

A total of 38 Raman spectra of human blood and 36 Raman spectra of human-unlike nonhuman blood (spectra from nine chicken-blood samples, nine duck-blood samples, 11 goose-blood samples, and seven dove-blood samples) were used to build model 1. A total of 109 Raman spectra of human blood (1–109) and 97 of nonhuman blood (samples from 110 to 187, which included chicken, duck, goose, and dove, combined with the samples from 188 to 206, which included pig and sheep) were used to validate the effectiveness of the model. The results are shown in Fig. 3. The predicted value for the samples ranging from 1 to 109 was <1.5 , which was discriminated as human blood by the PLS model. The predicted value for the samples ranging from 110 to 187 was >1.5 (except for the 149th sample, whose predicted value was 1.43), which was discriminated as nonhuman blood. However, for the samples ranging from 188 to 206, the results obtained by the PLS model were incorrect: the nonhuman blood was discriminated as human blood. We defined this blood as human-like nonhuman blood.

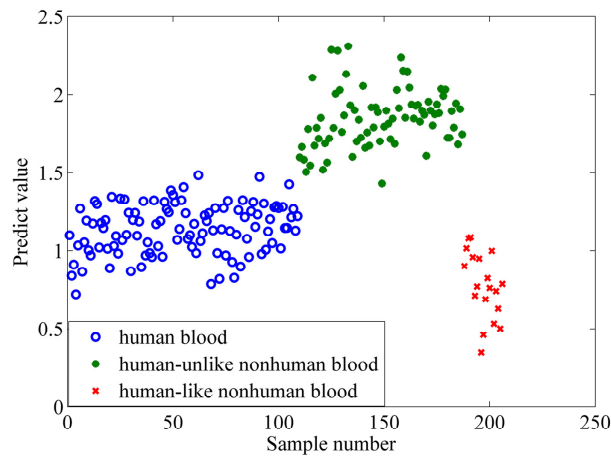


Fig. 3. Predicted value obtained by model 1.

The same 38 Raman spectra of human blood and 30 Raman spectra of human-like nonhuman blood (16 rabbit blood, eight pig blood, and six sheep blood) were used to build model 2. The human blood determined by model 1 was used to validate the model. The results are shown in Fig. 4. Among the 109 samples of human blood, only the 21st was discriminated as nonhuman blood, and among the 20 samples of nonhuman blood, the 149th, 113rd, and 114th were discriminated as human blood.

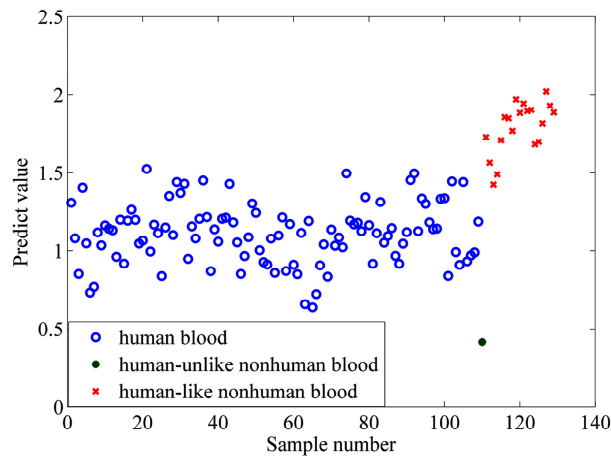


Fig. 4. Predicted value obtained by model 2.

Combining the results obtained by models 1 and 2, only one human blood sample and three nonhuman blood samples were mis-discriminated by the dual-model method. The final discrimination accuracy was 98.1%, and the TPR for human blood was 99.1%.

3.2 Blind test and external validation

Blind test and external validation [16] were performed to validate the performance of the model. A total of 12 spectra were randomly chosen from the above species for the blind test of the dual-model. Raman spectra of blood originating from a mouse and dog, which are not included in the above species, were adopted for external validation. The discrimination results of a blind test and external validation are shown in Fig. 5 and Table 1. The accuracy of the

blind test and external validation were 100%. These results illustrate the classification ability of the method for discriminating human blood.

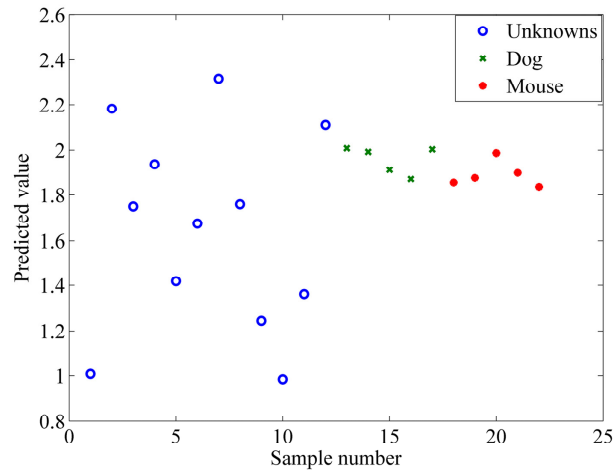


Fig. 5. Predicted value of the blind test and external validation.

Table 1. Blind Test and External Validation

Spectrum	Binary	Actual identify
Unknown 1	Human	Human
Unknown 2	Animal	Dove
Unknown 3	Animal	Pig
Unknown 4	Animal	Pig
Unknown 5	Human	Human
Unknown 6	Animal	Sheep
Unknown 7	Animal	Chicken
Unknown 8	Animal	Chicken
Unknown 9	Human	Human
Unknown 10	Human	Human
Unknown 11	Human	Human
Unknown 12	Animal	Duck
Mouse 1	Animal	Mouse
Mouse 2	Animal	Mouse
Mouse 3	Animal	Mouse
Mouse 4	Animal	Mouse
Mouse 5	Animal	Mouse
Dog 1	Animal	Dog
Dog 2	Animal	Dog
Dog 3	Animal	Dog
Dog 4	Animal	Dog
Dog 5	Animal	Dog

3.3 Comparison with single PLS model

The traditional PLS model was analyzed. A total of 38 Raman spectra of human blood, 36 Raman spectra of human-unlike nonhuman blood, and 30 Raman spectra of human-like nonhuman blood were used to build the model. All the Raman spectra used to build the model and those used to validate the model were the same as those used in the dual-model method, and results are shown in Fig. 6. Among the 109 human blood samples used to validate the model, 14 were discriminated as nonhuman blood, and among the 117 nonhuman blood samples, 11 were discriminated as human blood. The discrimination accuracy was 87.9%, and the TPR for discriminating human blood was only 87.2%.

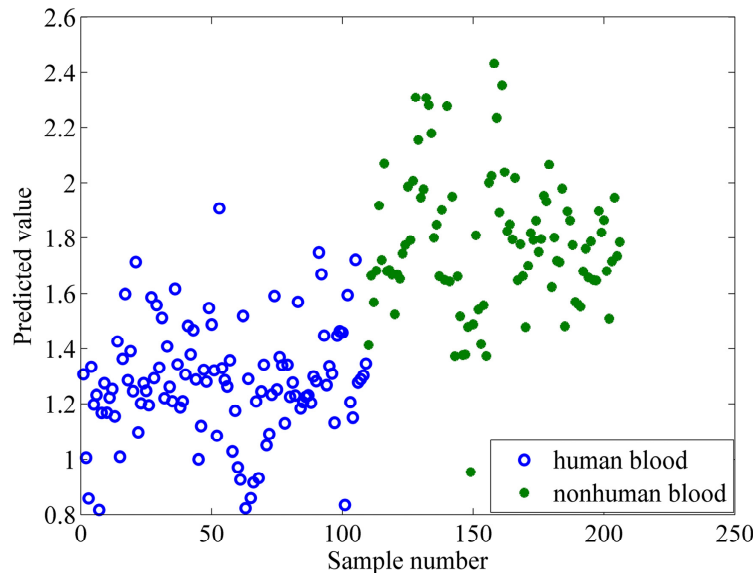


Fig. 6. Predicted value obtained by the traditional PLS model.

For customs inspection and forensic applications, discriminating human blood correctly is more important than accuracy discriminating between human and nonhuman blood. Table 2 presents the results of the dual-model and traditional PLS methods. The TPR for human blood reached 99.1% for the dual-model method, indicating an increase of 13.6% compared with the PLS method. All the parameters indicate the effectiveness of the proposed method for improving the accuracy of discrimination.

Table 2. Comparison of Dual-Model and Traditional Single-Model Methods

Methods	True positive rate	False positive rate	True negative rate	False negative rate
Dual model	99.1%	2.6%	97.4%	0.9%
Traditional single model	87.2%	9.4%	90.6%	12%

For both models, the values 1 and 2 represent human and nonhuman blood, respectively. To estimate the models for discriminating human blood, the receiver operating characteristic curve (ROC) was analyzed. To set the human blood as positive, the reciprocal of the predicted value was used. Finally, 0.5 (negative) was set to represent nonhuman blood, and 1 (positive) was set to represent human blood. In Fig. 7, the ROC curves of human blood versus nonhuman blood obtained using the dual-model and single-model methods are shown. The area under the ROC curve is one of the parameters that can be used to estimate the performance of the models. For the dual-model and traditional single-model analyses, the area

under the curve was 0.998 and 0.947, respectively. The results of the dual-model analysis are more promising.

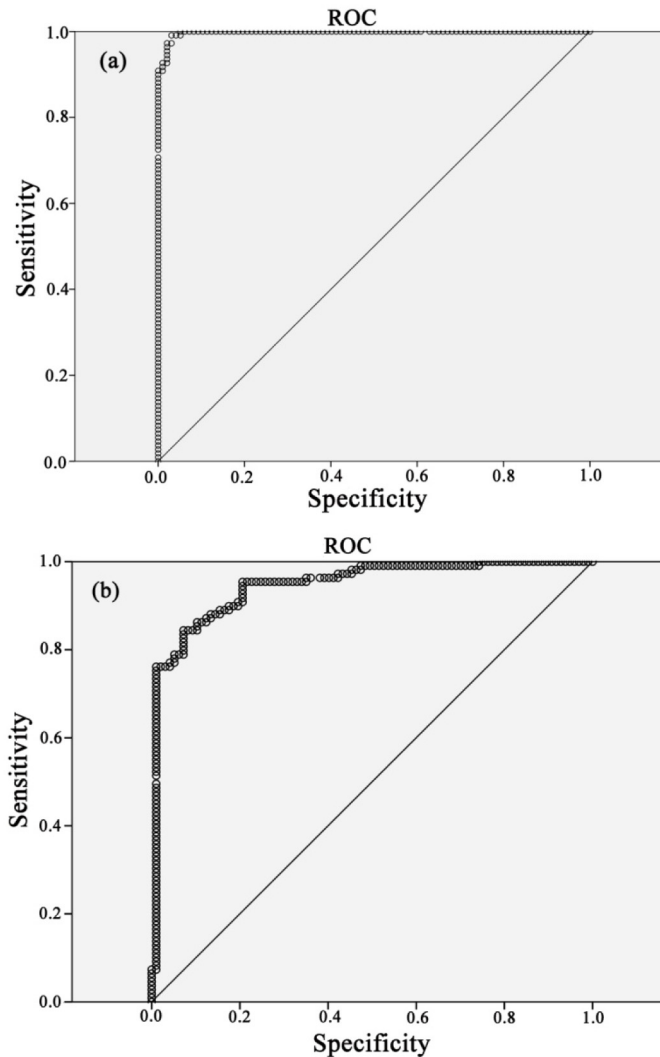


Fig. 7. ROC curves of human blood versus nonhuman blood. (a) Result obtained by the dual-model analysis method (area under the ROC curve is 0.998). (b) Result obtained by the traditional single-model analysis method (area under the ROC curve is 0.947).

4. Conclusion

We report a dual-model method for improving the accuracy and TPR of human-blood discrimination. In total, 332 Raman spectra from 10 species were used to demonstrate the proposed method. The results achieved were compared with those of the traditional PLS model. The TPR for human blood reached 99.1%, which is 13.6% higher than that for the PLS model, and the accuracy of discrimination of human and nonhuman blood reached 98.1%, which is 11.6% higher than that for the PLS model. The comparison results reveal that the dual-model method can improve the accuracy and TPR for discriminating human blood and is highly suitable for customs inspection and forensic applications. In a future work, the effects of the preservation time and conditions will be studied and discussed.

Funding

National High Technology Research and Development Program of China (863 Program 2015AA021105); Key project of Jiangsu Province (Grant No. BE2016090 and BE2016005-2); National Key Instrument Developing Project of China (Grant No. ZDYZ2013-1); State Key Project of China (Grant No. 2016YFB0402202); Opened Fund of the State Key Laboratory on Integrated Optoelectronics (grant number IOSKL2016KF12); Postdoctoral Funds of Jiangsu Province (Grant No 1701045B); ShenKang Developing Center (Grant No SHDC12017X10).

Disclosures

The authors declare that there are no conflicts of interest related to this article.